STATISTICS FOR MANAGEMENT (Session 1)

Zack Kertcher

IDS 570
FALL 2016

Plan

1. Why statistics?

2. Why R?

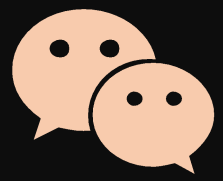
3. Syllabus and class logistics

Why Statistics?

What's statistics?

"…. collections of information of all types, [including] analysis and interpretation of such data… [using] statistical inference" (Wikipedia)

Examples of business applications using statistics



	Global	N. America	EMEA	APAC	L. America
Customer Relationship Management	1	4	1(t)	2(t)	4
Benchmarking	2(t)	2(t)	1(t)	14	2
Employee Engagement Surveys	2(t)	1	5	8	9(t)
Strategic Planning	2(t)	2(t)	9	5(t)	1
Outsourcing	5	6	3(t)	5(t)	9(t)
Balanced Scorecard	6(t)	7(t)	3(t)	15(t)	3
Mission and Vision Statements	6(t)	5	8	18	5
Supply Chain Management	8	7(t)	10	2(t)	13(t)
Change Management Programs	9	9	6(t)	21	9(t)
Customer Segmentation	10	14(t)	6(t)	12(t)	7
Core Competencies	11(t)	10	–	7	–
Big Data Analytics	11(t)	–	–	1	–
Total Quality Management	11(t)	–	–	4	–
Satisfaction and Loyalty Management	16	–	–	9	–
Digital Transformation	19(t)	–	–	10	–
Business Process Reengineering	15	–	–	–	6
Strategic Alliances	17	–	–	–	8

Note: (t)=tied

Source: Bain & Co. 2015 (+13,000 respondents in +70 countries)

In customer relationship management

Why R?

R

The 👎

`> command line`

`You need to code`

Although you'll need to write code in Excel (VBA) and any other statistical software

Initially high learning curve

r4stats.com/articles/why-r-is-hard-to-learn/

R

The 👍

Increasingly user friendly (RStudio!)

F R E E

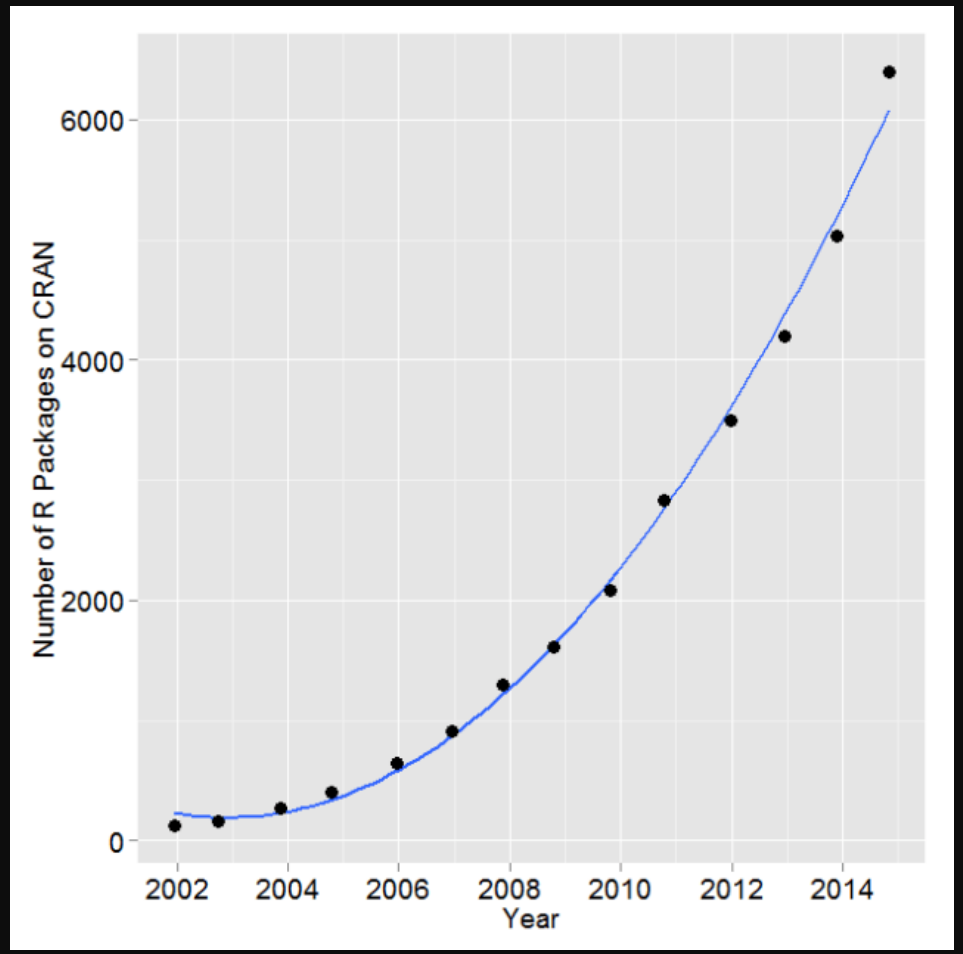
Excellent documents and many, many recipes
(R=550 blogs; Python=60; SAS=40)

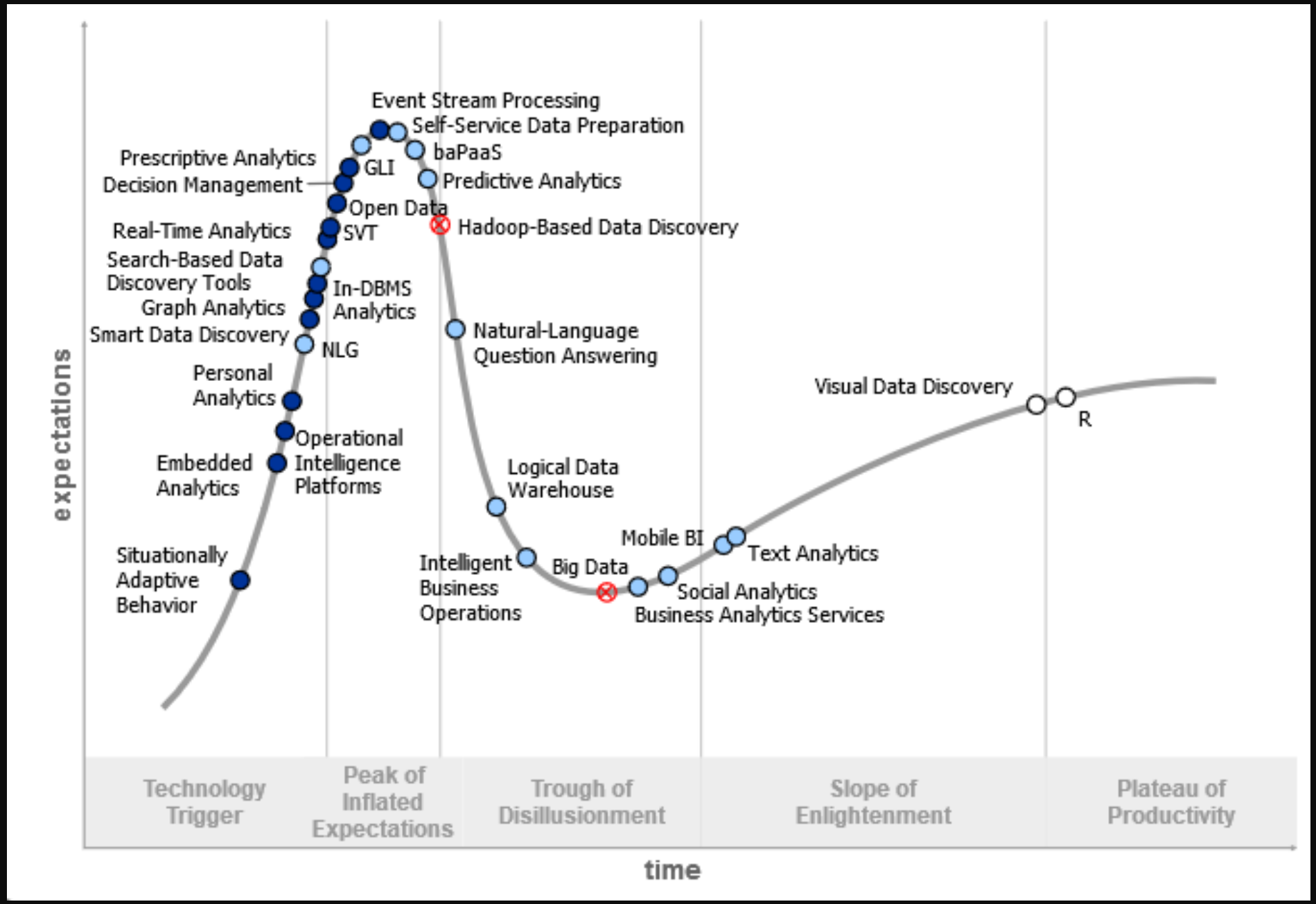
Incredible set of packages makes it extremely capable
(8,000+, and they are free, too)

R

"During 2015 alone, R added more functions/procs than SAS Institute has written in its entire history"

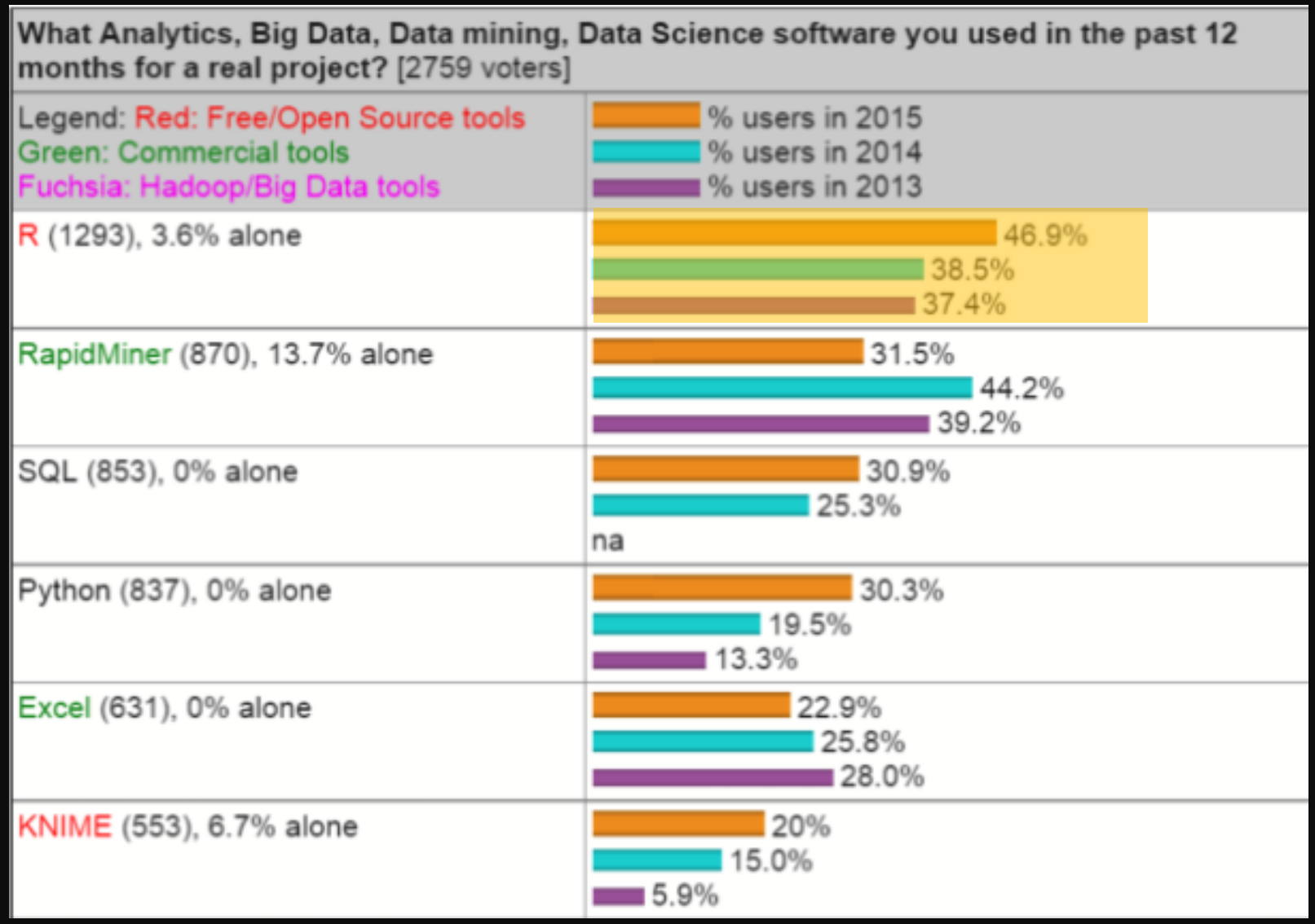
Source: r4stats





Source: Hype Cycle for Business Intelligence and Analytics, Gartner 2016

R



Source: KDnuggest 2015 (2,759 respondents)

R

SKILL	2015	YR/YR CHANGE
HANA (High Performance Analytical Appliance)	$ 154,749	n/a
Cassandra	$ 147,811	14.9%
Cloudera	$ 142,835	12.6%
PaaS (Platform as a Service)	$ 140,894	8.3%
OpenStack	$ 138,579	19.4%
CloudStack	$ 138,095	20.0%
Chef	$ 136,850	10.8%
Pig	$ 132,850	6.7%
MapReduce	$ 131,563	3.3%
Puppet	$ 131,121	9.2%
TcL (Tool Command Language)	$ 130,906	17.5%
Sqoop	$ 130,865	14.5%
NoSQL	$ 130,290	9.9%
Hive	$ 129,400	7.1%
Hadoop	$ 128,888	6.2%
UML (Unified Modeling Language)	$ 128,198	12.1%
SDN (Software Defined Network)	$ 127,464	12.0%
Omnigraffle	$ 127,392	11.1%
Fortran	$ 127,359	24.1%
SOA (Service Oriented Architecture)	$ 127,268	7.4%
R	$ 126,249	9.7%
Docker	$ 126,131	n/a
Netezza	$ 126,035	13.0%

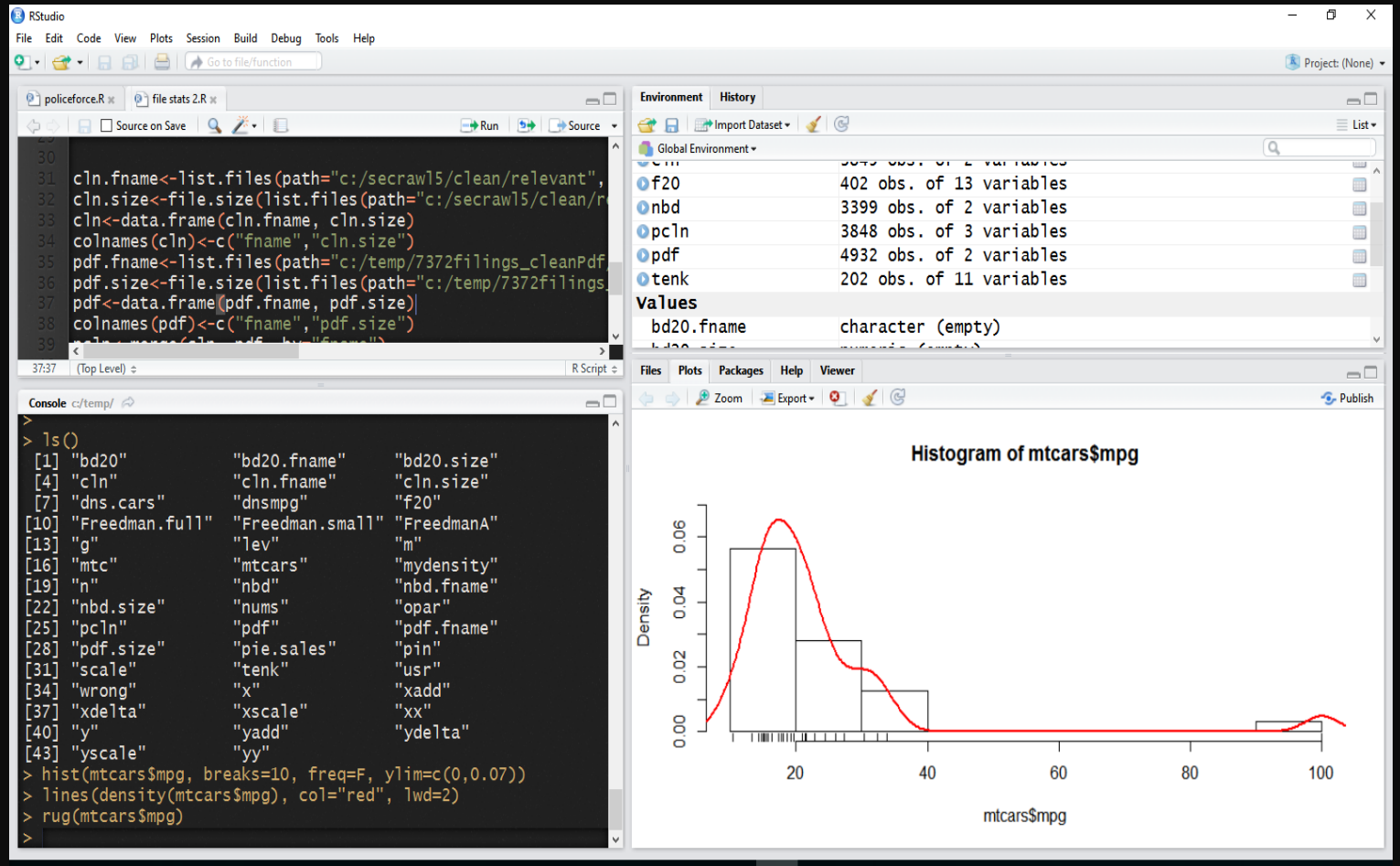
Source: Dice 2016
(+16,000 tech professionals)

Hello R (RStudio)!

R Basics (orientation)

Code /data

R console



R environment

Graphics

R Basics (orientation)

Know the shortcuts

Tab = autocomplete

CTRL (Command for Mac) +Up arrow = browse through a list of commands you've entered.

CTRL (Command for Mac) + Enter = copy current line, or multiple lines, from editor to console

R Basics (orientation)

- ✔ Create a folder called IDS570 under
- ✔ In RStudio, change your working directory to IDS570
- ✔ Install the "ggplot2" package

Solution

```
# you can create a folder by using your
computer's operating system, or through R:

getwd()
# [1] “c:/users/zack kertcher/documents"
dir.create("c:/users/zack
Kertcher/documents/IDS570")
setwd("IDS570")

install.packages('ggplot2')
```

R Basics (operations and assignment)

- ✔ Assign the numbers 4, 3.7, 4, 3.5 to object score
- ✔ Assign the number 4 to object tests
- ✔ Assign the result of score divided by tests to object avg.score
- ✔ Print the content of avg.score
- ✔ Assign the number 4 to object score
- ✔ Did avg.score change? Why?

Solution

```
score <- c(4,3.7,4,3.5)
tests <- 4
avg.score <- score/tests
avg.score
score <- 4
avg.score
# It's content did not change because the object
avg.score itself was unchanged. If we were to re-
run the assignment, it would have changed.
```

R Basics (data types)

- ✔ What is the data class of avg.score?
- ✔ Add the value "employee" to avg.score like this: avg.score <- c("employee", avg.score)
- ✔ Print the content of avg.score
- ✔ What is the class of avg.score now? Why is it different?

Solution

```
class(avg.score)
# [1] "numeric"
avg.score <- c("employee", avg.score)
avg.score class(avg.score)
# [1] "character"

# The data type has changed because we added a
character value to avg.score
```

R Basics (functions)

- ✔ Generate a sequence of numbers from 7 to 100, by 4.05, and assign it to my.seq
- ✔ View the last 4 numbers like this: tail(my.seq,n=4)
- ✔ Round the last four numbers of my.seq to a single decimal, and assign them to new.seq

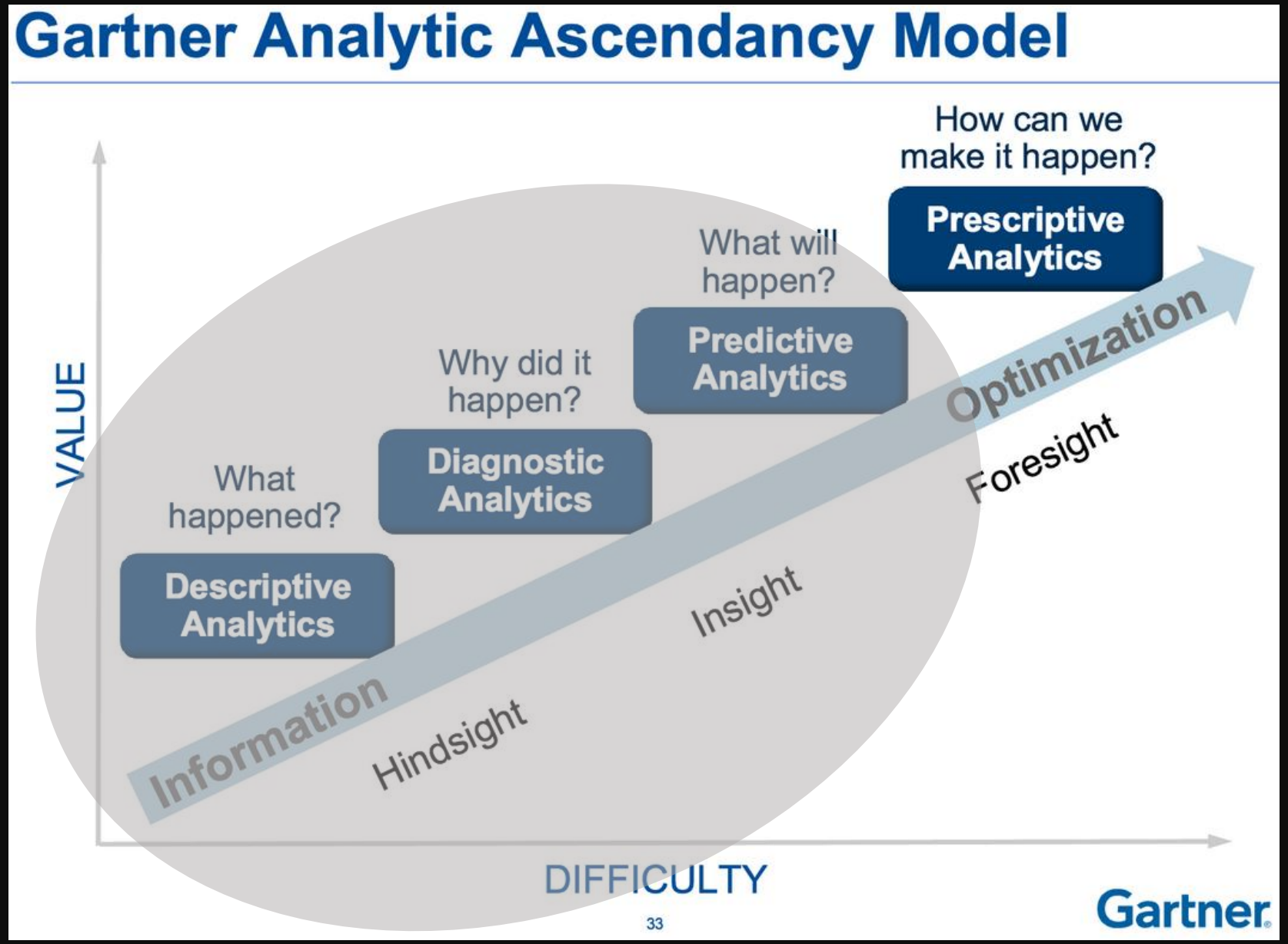
Solution

```
my.seq <- seq(from=7, to=100, by=4.05)

new.seq <- round(tail(my.seq,n=4),1)
```

Course syllabus and logistics

- ✔ Reading? None. Recommended reading posted on Blackboard.
- ✔ Questions? Office hours: after every class or by appointment.
- ✔ Coursework
  - Homework: Most weeks, short. Submit by the following class, and two in-class quizzes (25%)
  - Team project: presentation + report (25%)
  - Exams: midterm and final (50%)
- ✔ Feedback? www.admonymous.com/ids_570



Source: Gartner 2015

Class	Date	Topic
1	8/27	Introduction to Statistics and R
-	9/3	No Class (Labor Day)
2	9/10	Data
3	9/17	Descriptive statistics
4	9/24	Probability and univariate distributions
5	10/1	Bivariate associations
6	10/8	Tables and plots
7	**10/15**	**Midterm Exam**
8	10/22	Statistical research design
9	10/29	Hypothesis testing
10	11/5	Analysis of variance
11	11/5	Linear regression
12	11/12	Advanced topics
13	**11/19**	**Project presentations**
-	11/26	No class (Thanksgiving)
14	12/03	Exam Review
15	**12/10**	**Final exam**

Demo and Sneak Preview to Data

<>